

NIH "Fast Track" Directory Project Definition

DRAFT 11

Tuesday, June 29, 1999



Table of Contents

1.0	Background.....	1
2.0	Purpose	1
3.0	Project Definition.....	2
3.1	System Components.....	2
3.2	Enterprise Databases and Directories Involved.....	2
3.3	Directory Relational Database	4
3.3.1	Strategy for FTRDB Initial Load	5
3.3.2	Plan for FTRDB Maintenance	5
3.3.3	Interface to FTRDB	6
3.3.4	Personal Identification Numbers.....	6
3.4	Record Linking Engine	6
3.5	Directory Registration and Update Service.....	7
3.5.1	Functionality	7
3.5.2	User Interface.....	7
3.6	Interface to PAID.....	8
3.7	Interface to Telecommunications Database	8
3.8	Interface to NIH Email Directory and Forwarding Service	8
3.9	LDAP Directory Server	8
3.10	Security Requirements.....	9
	“Fast Track” Database Entity-Relation Diagram.....	11
Appendix B	“Fast Track” Attribute Descriptions.....	17
Appendix C	“Fast Track” Database Creation.....	23
C.1	Record Linking	23
C.2	Attribute Selection	24
Appendix D	Probabilistic Record Linkage.....	27
D.1	Record Linking Example	28
D.2	Record Linking References.....	28
Appendix E	“Fast Track” Update Service User Interface	31
Appendix F	Outstanding Questions and Issues.....	33
Appendix G	Wish List.....	35

List of Tables

Table 3-1 Enterprise Databases and Directories 3
Table 3-2 NIH Databases Containing Individual Identifying Information 4
Table B-1 Abbreviations for Data Sources 17
Table B-2 Private Individual Identifying Information Associated with NIH UIDs 17
Table B-3 Private Home and Personal Locator Attributes 18
Table B-4 Public Labeling Attributes 19
Table B-5 Public Organizational Attributes 20
Table B-6 Public Locator Attributes 20
Table B-7 Security Attributes 22
Table B-8 Ancillary Attributes 22
Table D-1 Example Attribute Comparison Outcome Frequencies 28

1.0 Background

As a result of the NIH Director's Retreat of September, 1996, the NIH Director commissioned an NIH Information Technology Central Committee (ITCC) to make recommendations for improving IT management at NIH. Among its seven major recommendations, the ITCC's report of November, 1996 included the following:

- Develop a strategy for a secure networking environment that includes the Internet.
- Develop a strategy for a secure, centrally coordinated NIH electronic directory that logically coordinates directories for e-mail, personnel, parking, etc., and fully implements de-registration activities.

The NIH Director gave the acting NIH CIO the task of implementing these recommendations, who in turn commissioned the NIH Architectural Management Group (AMG), which is comprised of representatives from each NIH ICD, to undertake this work. The AMG's *Report on Interoperability at the NIH* issued in May, 1997 made the following recommendations relating to the security and directory strategies:

- Establishment of the NIH centrally supported electronic directory is a critical priority. [...] Action on recommendations for the central electronic directory and for a secure networking environment is paramount. If consolidated and centrally coordinated actions are not taken immediately, individual institutes will provide their own solutions, which be more expensive to maintain and will complicate interoperability issues. Further, development and implementation of the directory prerequisite to the emplacement of network security at the NIH. In addition, the directory must be recognized by all ICDs as authoritative source for directory information.
- Establish a central authentication service that uses public key or shared secret key security standards and provides electronic credentials for all the NIH workforce. The credentials issued by the central NIH authentication service will carry a unique personal identifier assigned to each NIH employee. To enable these credentials to grant access to computing services, the unique personal identifiers will need to be associated with and used consistently across various subsidiary certificate authorities, NOS security registries, and directory services.
- Unique personal identifiers (not the SSN) must be defined. This will allow integration with systems based on relational databases. Declare directory presence a prerequisite for NIH services. This will help ensure the accuracy and currency of the information because presence in the directory will be required before other NIH resources can be used.

2.0 Purpose

The purpose of the "Fast Track" directory is to quickly bring up a working, but limited, directory containing NIH "white pages" information. The motivation for this effort is:

- Lightweight Directory Access Protocol (LDAP) and meta-directory technology and products are immature. These are needed to meet the full set of directory requirements as identified by the AMG Technical Subcommittee (TSC). Experts expect that critical LDAP security and other features will not be standardized and generally available for at least 2 years, and at least several months will be required just to evaluate and select a meta-directory product.
- Implementing and distributing Unique Identifiers (UIDs) and corresponding Personal Identification Numbers (PINs) to all individuals who use NIH services and facilities, and creating and updating the NIH meta-directory will have a significant impact on existing NIH business processes and

applications. Reengineering these requires a long-term effort, involving the owners of the affected business processes and applications.

- Existing NIH directories and databases are inconsistent, inaccurate, and incomplete to varying extents. We do not know how much effort will be required to clean and merge these in order to create the NIH meta-directory, nor how complete and accurate the meta-directory must be in order to support anticipated uses, particularly those related to security, nor what level of completeness and accuracy can be practically achieved.

Building the "Fast Track" directory will give us valuable experience with:

- cleaning and merging some of NIH's major enterprise databases and directories
- operational use of a directory for some existing applications, such as ITAS
- key NIH business processes and owners, such as ID badge

3.0 Project Definition

In order to shorten the development time for the "fast track" directory, we must:

- Eliminate dependencies on LDAP security features that are not yet available
- Avoid use of meta-directory products
- Eliminate or minimize impact on existing business processes and applications

We can accomplish this by adopting the following design rules and limitations:

- Provide read-only access to LDAP clients
- Provide only public information to LDAP clients
- Integrate only a minimal number of enterprise databases and directories
- Loosely couple to ID badge and telephone directory systems via mechanisms such as paper forms or email
- Limit the directory to contain only person objects
- Limit the directory population to those individuals for whom administrative officers (AOs) serve as registration officials
- Limit update of directory information to only administrative officers
- Limit organizational information to that obtainable from the HRDB, or the NIH Scientific Directory, when this becomes available

3.1 System Components

The major components of the "fast track" directory are:

- A relational database containing (1) NIH UIDs and associated individual identifying information, (2) public "white pages" information that is published in the directory, and (3) other private information, which might be made available in the directory when it can be adequately secured
- A record linking engine
- A Web-based directory registration and update service
- An ITAS-based authentication service
- A read-only LDAP directory server

3.2 Enterprise Databases and Directories Involved

Table 3-1 summarizes the existing enterprise databases and directories which will either provide information to build the "fast track" directory database, or will interact with the "fast track" directory update service.

Table 3-1 Enterprise Databases and Directories

<i>Name</i>	<i>Population Description</i>	<i>Pop. Size</i>	<i>Data Quality</i>	<i>Use</i>
Human Resources DB (HRDB)	All NIH FTE employees (Federal Civil and Public Health Service) at all NIH sites	18,000	Very good	Initial load, compare to directory and produce exception reports ¹
Fellowship Payment System (FPS)	All NIH non-FTE employees (Visiting Fellows and IRTAs) at all NIH sites	2,800 active <ul style="list-style-type: none"> • 1,200 Visiting Fellows • 1,600 IRTAs 	Good	Initial load
J.E. Fogarty International Center (JEFIC) DB	All NIH foreign visiting scientists at all NIH sites	2,100 active <ul style="list-style-type: none"> • 1,200 Visiting Fellows • 500 Visiting Scientists & Associates • 400 other 	Good	Initial load
Parking and ID Badge (PAID) DB	All individuals working at NIH sites in Maryland, except FDA employees who do not have NIH parking permits or participate in TRANSHARE	32,000 <ul style="list-style-type: none"> • 24,000 employees • 6,000 contractors • 1,500 volunteers • 400 guests 	Fair - Good	Initial load, receive directory updates from Aos
Telecommunications DB	Permanent NIH Federal employees, Temporary FTE employees >1 year, Temporary Federal physicians >6 months, and other non-Federal employees (40% of records, no SSN), at all NIH sites	17,500	Fair	Initial load, receive directory updates from Aos
NIH Email Directory and Forwarding Service (PH)	Most individuals registered for one or more NIH email services	>29,000	Poor	Initial load
Integrated Time and Attendance System (ITAS)	All NIH FTE employees (Federal Civil and Public Health Service) at all NIH sites	18,000	Very good	Authentication of AOs by directory registration and update service ²

¹Note: (1) there can be a two week to 1 month delay from the time an employee begins or terminates employment before the HRDB is updated, and (2) information in the FTRDB will more likely be current, since it is operational, unlike the official information maintained in the HRDB (e.g., organizational information).

² ITAS should also be able to use the "fast track" directory service to determine an employee's email address from their SSN, and later to perform authentication with an employee's UID and PIN instead of their SSN and ITAS-specific password.

Of particular importance is *individual identifying information*, which is associated with an NIH UID so that an individual is assigned the same NIH UID each time they enter the NIH workforce. Table 3-2 summarizes the individual identifying information available in existing NIH databases.

Table 3-2 NIH Databases Containing Individual Identifying Information

<i>Information</i>	<i>HRDB</i>	<i>JEFIC</i>	<i>FPS</i>	<i>PAID</i>	<i>Telecom</i>	<i>ITAS</i> ³
SSN	X		X		X	X
Date of Birth	X	X				
Place of Birth		X				
Sex	X	X				
Home Address	X		X	X		
Foreign Home Address		X				
Home Telephone					X ⁴	

3.3 Directory Relational Database

The "fast track" directory relational database (FTRDB) will be implemented with Oracle on CIT's Digital Alpha Enterprise Open System; however, it will be accessed via the ODBC/SQL standards such that it could be readily moved to a different database product or platform. The tables (See Appendix A) store:

- NIH UIDs and associated individual identifying information
- public "white pages" information that is published in the LDAP directory
- other private information that is not published in the LDAP directory

Access to individual identifying information covered by the Privacy Act will be controlled by views, and the NIH UIDs of individuals accessing the attributes listed in "Table B-2 Private Individual Identifying Information Associated with NIH UIDs" will be logged, along with the time of access.

The FTRDB will also contain:

- tables for validating data entry of:
 - job series,
 - NIH organization names and corresponding Standard Administration Codes (SACs),
 - organizational abbreviations used for the NIH Telephone and Service Directory,
 - country designations,
 - name components: personal title, generation qualifier,
 - building designations and addresses, and
 - work telephone area codes and exchanges;
- tables to hold working copies of other databases and directories, together with their linked UIDs, weights indicating the reliability of the linkage, the time at which the linkage was made, and the last modification time of the linked record⁵;
- tables for mapping NIH organization and building designations from the codes used by HRDB, FPS, JEFIC, PAID, and Telecommunications to the designations used by the FTRDB and directory;
- a table to store given names and their corresponding nick names;
- tables to store weights for use by the record linking engine;
- a modification log to record the identity of the modifying agent, the time of the change, and the old and new values.

³ ITAS contains the same individual identifying information as the HRDB.

⁴ For Clinical Center on-call staff

⁵ Records in the Telecommunications Database do not have last modification timestamps. We are investigating the possibility of calculating crude last modification dates by comparing archived copies of the database.

See Appendix B for a detailed description of the main FTRDB fields. While the FTRDB data dictionary will include all `.nihInetOrgPerson` attributes, the only attributes that will actually have values will be those that (1) can be initially loaded from one of the databases listed Table B-1 or (2) can be entered and updated by AOs via the directory registration and update service.

3.3.1 Strategy for FTRDB Initial Load

The process of building the FTRDB will involve the following operations:

- **Data element analysis**—surveying the values of a field or attribute in a database or directory to determine what information is held and in what formats. Particularly needed for format-free values such as found in the NIH Email Directory.
- **Parsing**—associating data values with data types; for example, associating a data value such as “Dr. Jane Ann Doe” with the types `personalTitle`, `givenName`, `middleName`, and `surname`, respectively.
- **Standardization**—transforming field or attribute values to a standard format so they can be represented consistently in the NIH directory and so they can be easily compared; for example, converting “Natcher Building” to its building number, “45”.
- **Blocking (or clustering)**—grouping the records in a database by some criteria, such as the Soundex or NYSIIS phonetic encoding of surname, in order to reduce the number of possible records for linking to a tractable number.
- **Matching**—comparing two values to determine if they agree, partially agree, or disagree.
- **Record Linking**—determining if two records identify the same individual.
- **Attribute selection**—deciding which values from a set of linked records to choose to load into attributes in the NIH Directory. This arises when the same attribute represented in two or more linked records has two or more different values.

The strategy for loading the FTRDB is:

1. Deduplicate all databases/directories used for the initial load.
2. Use the best data (in HRDB, FPS, and JEFIC) and the best keys (the SSN and JEFIC case number) to create entries with UIDs associated with quality individual identifying information.
3. Use the SSN to link to records from ITAS and the Telecommunications databases in order to load in work locator attributes. These provide more common attributes on which to base links to the other databases, which do not contain the SSN.
4. Link records from PAID on name and other common attributes, loading the additional attributes it contains.
5. Add new UIDs for active records in the PAID database that have not yet been associated with a UID. These “weak” UIDs will have little or no associated individual identifying information.
6. Link records from the Telecommunications and Email directories on name and other common attributes, loading the `@nih.gov` email address and other additional attributes they contain.

At each stage, conflicting attribute values for the same individual may be found in different databases, in which case values will be selected as described in Appendix C.

A detailed plan for loading the FTRDB is described in Appendix C.

3.3.2 Plan for FTRDB Maintenance

Once the FTRDB is built, subsequent addition, update, and deactivation of records will be done by AOs by means of the directory registration and update service. However, viewed as a replacement for the current *Request for DHHS Identification Card* (form NIH 1308-4/5) and *Request to Change NIH Directory Information* (form NIH 433) business processes, this procedure will not track 100% of the directory population. First, the following groups of individuals are not issued NIH ID badges:

- Employees working at sites other than NIH sites in Maryland⁶
- FDA employees working at NIH sites in Maryland⁷

Second, temporary FTE employees, non-FTE employees, contractors, volunteers, and guests are not supposed to be listed in the *NIH Telephone and Services Directory*. Thus, there are populations of individuals not covered by either business process, as currently defined. But it is hoped that AOs will be willing to maintain directory information for these additional groups.

It will thus be necessary to periodically update the FTRDB with changes made to the HRDB, FPS, JEFIC, PAID⁸, and Email systems that have bypassed the directory registration and update service. [The method for performing such updates needs to be better defined.]

3.3.3 Interface to FTRDB

[More detail to be supplied by Bob on creating new UIDs/entries and interface to the HRDB]

3.3.4 Personal Identification Numbers

A secret 4 – 8 digit Personal Identification Number (PIN) , perhaps derived from an individual's SSN, date of birth, or voice mail PIN, will be associated with each UID by storing it in the `userPassword` attribute. The "fast track" update service (see Section 3.5) will print an individual's PIN, along with instructions for protection and use, on paper for the registering AO to give to a new employee or contractor. To protect from loss or theft, the paper will not contain any identification of the owning individual. In future phases of the directory project, an individual will be able to use the secret PIN together with their UID to authenticate to automated systems.

3.4 Record Linking Engine

As noted previously, *record linking* refers to the process of determining if two records belong to the same individual. Record linking has several uses in connection with the NIH directory:

1. elimination of multiple records belonging to the same individual in existing databases,
2. linking records across two databases (e.g. the CSO Email Directory and the HRDB) so they can be associated with the same NIH UID,
3. determination of which UID, if any, belongs to a "new" employee who may have previously been assigned a UID, so the same UID can be reassigned, and
4. in future phases, joining newly-created entries from connected directories into the NIH meta-directory.

Record linking is easy in situations where a decision can be made based on the agreement or disagreement of a single attribute, for example, the SSN. However, it becomes more difficult when the records to be linked do not contain such an attribute, and the decision must be based either on a single attribute that may partially agree (such as a name) or several attributes of which only *some* may agree (such as organization, office address, and office telephone number).

The more difficult cases may be handled by applying *probabilistic record linking*, described in more detail in Appendix D. Briefly, the record linking engine calculates a number, called a *binit weight*, which is the \log_2 of the odds that two records constitute a *linked pair*, i.e., that they belong to the same individual. Thus, a positive binit weight of, say, +10 indicates that the odds are about 1,000 to 1 in favor of a linkage, a negative binit weight of -10 indicates odds of about 1,000 to 1 *against* a linkage (an *unlinked pair*), and a binit weight of 0 indicates even odds in favor of (or against) a linkage. Depending on the acceptable

⁶ Does RTP also use form NIH 1308-4/5??

⁷ Approximately 750 FDA employees with NIH parking permits or who participate in TRANSHARE **do** have records in the PAID database.

⁸ Is it possible to arrange for all changes to PAID to be made via the FTUS?

number of false positive and false negative links, and the number of borderline pairs one is willing to manually review, an upper and lower threshold can be established. Binit weights above the upper threshold are accepted as linked pairs, those below the lower threshold are accepted as unlinked pairs, and those between are subjected to manual review, perhaps suggesting additional tests to be incorporated in the linking engine to improve its discriminating power.

3.5 Directory Registration and Update Service

The "fast track" directory update service (FTUS) will enable AOs to register, update, and de-register NIH employees and contractors. It will be implemented as a web application server that will interact with AOs via Netscape or Microsoft browser clients and HTTP/HTML, and with the FTRDB via ODBC/SQL. Workflow procedures will be incorporated to allow AOs to delegate some duties to others with final approval being made by the AOs (for example, via email). The HTTP/HTML browser client will include a trusted certificate authority certificate which will enable a SSLv2 connection to be made to the FTUS. AOs will supply their SSN and ITAS passwords to the FTUS over this secured connection, which will query the ITAS database to validate passwords.

3.5.1 Functionality

The FTUS will allow an authorized AO to:

- Register a new employee or contractor. The AO will enter individual identifying information (see Table B-2 Private Individual Identifying Information Associated with NIH UIDs), and the FTUS will use the linking engine to determine if the individual has already been assigned a UID; if not, the FTUS will obtain a new UID. The AO will then enter or update the *Initial Source* or *Update From* fields marked "A" in the tables shown in Appendix B, and the FTUS will print the individual's PIN (see Section 3.3.4) for the AO to give to the new employee/contractor, and forward the new employee information to the Badge Office and Telephone Directory Unit,.
- Update the FTRDB information for an employee or contractor. The FTUS will also send the changes to the Telephone Directory Unit and the ID Badge/Parking Office.
- Deregister (deactivate) an employee or contractor upon termination of employment (separation). The FTUS will mark the individual inactive in the FTRDB, causing the individual's information to be removed from the LDAP directory, and also notify the Badge Office and Telephone Directory Unit.

The Badge Office and Telephone Directory Unit can also be authorized to use the FTUS to update the FTRDB with information received from walk-ins and paper forms. Updates made in this fashion will cause notification to be sent via email to the requesting AO and affected individual. Update access will be permitted to only those attributes present on the current *Request for DHHS Identification Card* (form NIH 1308-4/5) and *Request to Change NIH Directory Information* (form NIH 433).

The FTUS will be designed to allow ICs to easily extend it to collect additional information, automatically create LAN or email accounts, or send additional notifications, for example.

During the "Fast Track" phase, NIH UIDs will not be widely distributed, and individuals will in general not know their UID. We particularly need to provide email account administrators and others with a tool they can use to find someone's UID so they can begin to add the UID to the accounts they manage. Therefore, the FTUS will also provide an interface to allow anyone to search the *active* records in the FTRDB (using the linking engine) for an individual's UID by entering the individual's surname (*sn*) and any or all of the following public attributes: *cn*, *givenName*, *nihNickname*, *middleName*, *o*, *ou*, *nihCompanyName*, *telephoneNumber*, *buildingName*, *roomNumber*.

3.5.2 User Interface

Prototype FTUS user interface screens are depicted in Appendix E. These generally adhere to the following guidelines:

1. The collection of fields that appear together on a screen or tab correspond to the attributes described in the tables in Appendix B.
2. Fields for *required* attributes are visually distinguished from those for *optional* attributes.
3. Fields for *multi-valued* attributes are drop-down or scrollable fill-in lists.
4. Fields for attributes that a user is not authorized to *read* are completely filled in with “*”s.
5. Fields for attributes that a user is authorized to read, but not modify, are visually distinguished from those that a user is authorized to change.
6. All fields are fixed format, unless explicitly indicated otherwise in Appendix B.
7. Fields for attributes with FTRDB validation tables (job series, organization names, building names, work telephone area codes and exchanges, job series) are drop down pick lists.
8. Attributes listed in “Table B-2 Private Individual Identifying Information Associated with NIH UIDs” will not be displayed unless (a) the user is authorized, and (b) the user has indicated via a check box field that they need to see this information.

3.6 Interface to PAID

[To be supplied by Denney and Diane]

3.7 Interface to Telecommunications Database

[To be supplied by Dave]

3.8 Interface to NIH Email Directory and Forwarding Service

An AO will generally not know an individual’s @nih.gov email alias to enter into the FTRDB via the FTUS. Ideally, all email administrators will add the NIH UID to the email accounts they manage, and include the UID in the information they feed to the NIH Email Directory and Forwarding Service (PH). This would enable PH (or its replacement) to easily recognize duplicate entries, link entries and exchange attribute information with the FTRDB, and handle deregistration. Unfortunately, this is a difficult process to implement for the “Fast Track” because users will generally not know their UIDs⁹, and there are 23 email systems that feed PH, each with its own group of administrators.

The plan for dealing with this situation is:

- Add an NIH UID attribute to the existing PH system.
- Provide a tool that email administrators can use to look up an individual’s UID using public information, as described in Section 3.5.
- Identify the administrators responsible for the largest NIH email systems, and work with them to incorporate NIH UIDs.

[Note: Not a very convincing plan.]

3.9 LDAP Directory Server

An LDAP directory server, such as Netscape's, will provide read-only access to the active and listed public information contained in the FTRDB. This will be accomplished by transferring daily a copy of this information from the FTRDB to the LDAP directory server.

⁹ Users cannot generally be expected to know their NIH UID until completion of the “OneID” phase, after which all NIH ID badges will have been replaced with badges showing the UID.

3.10 Security Requirements

The following information is from the DHHS handbook located at <http://www.oir.nih.gov/policy/aissp.html>.

The Central Directory would have a sensitivity level of either 2 or 3, depending on exactly what data is included in the database.

C. Security Level Requirements

1. Level 1 Requirements

The controls required to adequately safeguard a Level 1 AIS, AIS facility, or ITU are those which would normally be considered good management practice. These include, but are not limited to:

- a) An employee AIS security awareness and training program.
- b) The assignment of sensitivity designations to every employee position.
- c) Physical access controls.
- d) A complete set of AIS documentation.

2. Level 2 Requirements

The controls required to adequately safeguard a Level 2 AIS, AIS facility, or ITU include all of the requirements for Level 1, plus the following requirements:

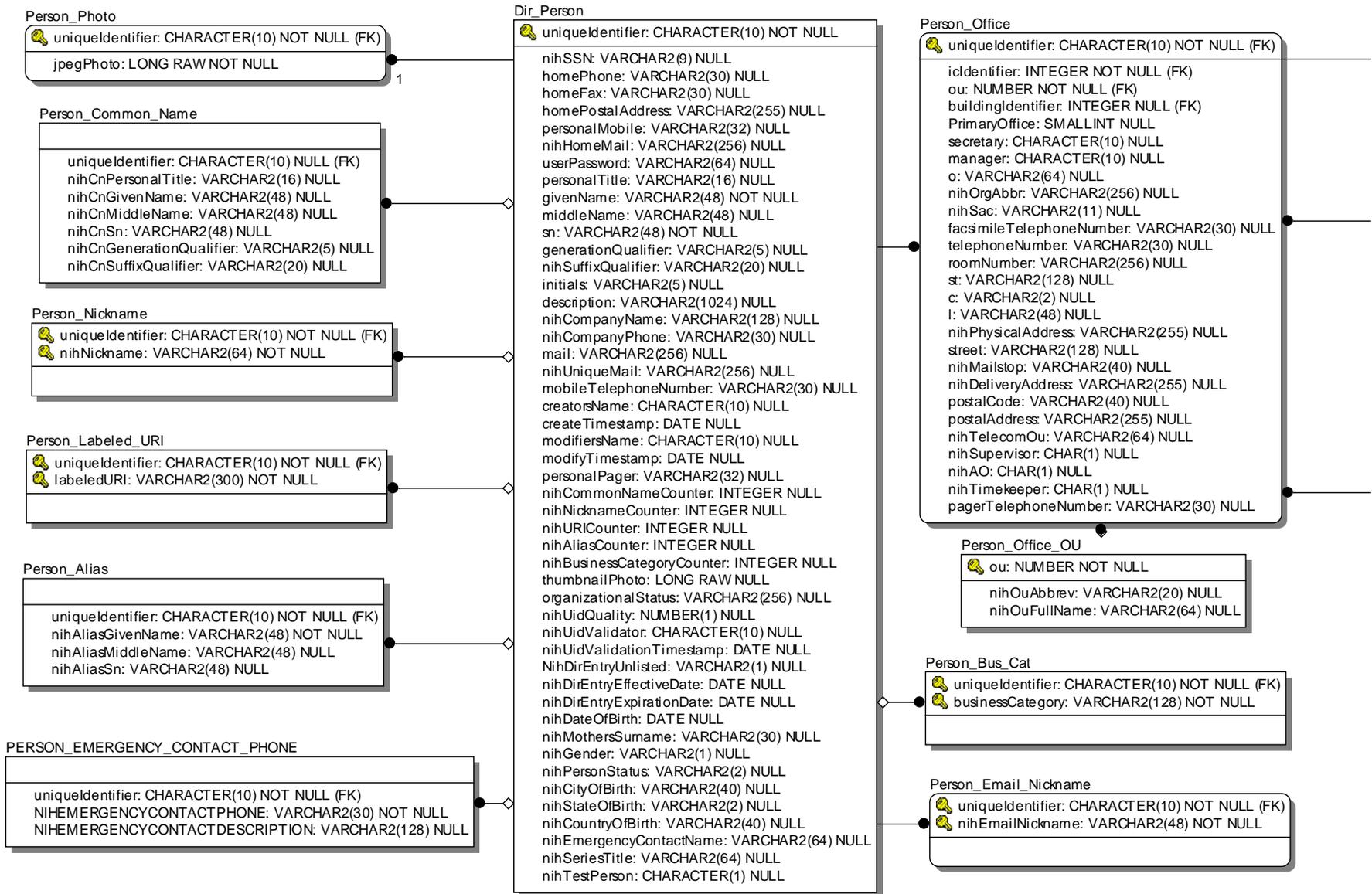
- a) A detailed risk management program.
- b) A CSSP for systems processing sensitive information.
- c) Record retention procedures.
- d) A list of authorized users.
- e) Security review and certification procedures.
- f) Required background investigations for all employees.
- g) Required background investigations for all contractor personnel.
- h) A detailed fire emergency plan.
- i) A formal written contingency plan.
- j) A formal risk analysis.
- k) An automated audit trail.
- l) Authorized access and control procedures.
- m) Secure physical transportation procedures.
- n) Secure telecommunications.
- o) An emergency power program.

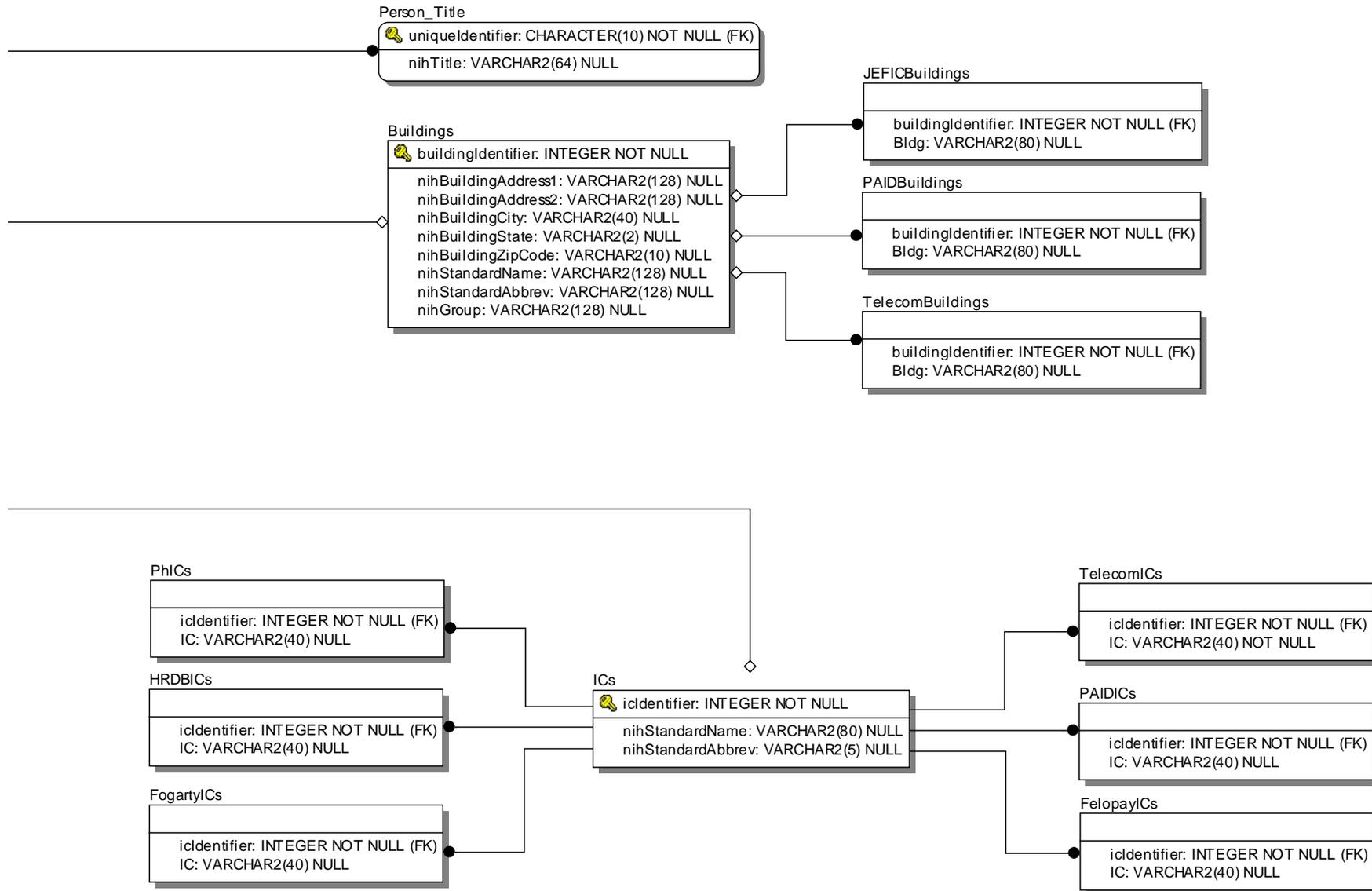
3. Level 3 Requirements

The controls required to adequately safeguard a Level 3 AIS, AIS facility, or ITU include all of the requirements for Levels 1 and 2, plus the requirement for an inventory of hardware and software.

CIT will need to develop a complete security plan for the directory before it becomes operational, in which all of these issues will be addressed.

Appendix A "Fast Track" Database Entity-Relation Diagram





HRDB

ICD: CHAR(6) NULL
 EMP_LNAME: CHAR(20) NULL
 EMP_FNAME: CHAR(11) NULL
 EMP_MI: CHAR(1) NULL
 SER: NUMBER(5) NULL
 PAY_PER_END_DTE: DATE NOT NULL
 POS_TITLE: CHAR(27) NULL
 SSN: VARCHAR2(9) NULL
 DOB: DATE NULL
 SEX: CHAR(1) NULL
 EMP_CITY_ST: CHAR(25) NULL
 ADM_CD: CHAR(11) NULL
 EMP_STR_ADD: CHAR(30) NULL
 EMP_ZIP_CD: VARCHAR2(5) NULL
 BINIT_WT: NUMBER NULL
 LINKTIMESTAMP: DATE NULL
 P_MIDDLENAME: VARCHAR2(30) NULL
 P_SN: VARCHAR2(30) NULL
 P_GENERATIONQUALIFIER: VARCHAR2(5) NULL
 IGNORE_ENTRY: CHAR(1) NULL
 P_GIVENNAME: VARCHAR2(30) NULL
 UNIQUEIDENTIFIER: CHAR(10) NULL
 P_NIHHOMEHOUSENO: VARCHAR2(6) NULL
 P_NIHHOMESTREETPREFIX: VARCHAR2(5) NULL
 P_NIHHOMESTREETNAME: VARCHAR2(20) NULL
 P_NIHHOME2NDSTREETNAME: VARCHAR2(10) NULL
 P_NIHHOMESTRUCTDESC: VARCHAR2(4) NULL
 P_NIHHOMESTRUCTID: VARCHAR2(5) NULL
 P_NIHHOMERRDESC: VARCHAR2(3) NULL
 P_NIHHOMERRID: VARCHAR2(5) NULL
 P_NIHHOMEBOXDESC: VARCHAR2(4) NULL
 P_NIHHOMEBOXID: VARCHAR2(5) NULL

SURNAMES

SURNAMEIDENTIFIER: NUMBER NOT NULL
 SURNAME: VARCHAR2(20) NULL

SURNAME_COUNTS

SN: VARCHAR2(48) NOT NULL
 COUNT: INTEGER NOT NULL

Surname_Prefixes

Prefix: VARCHAR2(20) NOT NULL

Generation_Qualifiers

Qualifier: VARCHAR2(20) NOT NULL

Personal_Titles

Title: VARCHAR2(20) NOT NULL
 Gender: VARCHAR2(20) NULL

GIVEN_NAMES

NAME: VARCHAR2(20) NOT NULL
 GENDER: VARCHAR2(1) NULL
 M_FREQ: NUMBER NULL
 F_FREQ: NUMBER NULL

CANS

ICD: CHAR(6) NOT NULL
 CAN: CHAR(7) NOT NULL

JEFIC_OrgStatus

CAT: VARCHAR2(2) NOT NULL
 organizationalStatus: VARCHAR2(1) NULL

SAC_CODES

ICD: CHAR(6) NULL
 ADM_CD: CHAR(11) NULL
 ADM_TITLE: CHAR(67) NULL
 ORG_INITS: CHAR(15) NULL
 DATE_ENTERED: DATE NULL
 UPDATE_CODE: CHAR(1) NULL
 STATUS: CHAR(1) NULL
 PGM_TYPE: CHAR(1) NULL
 CNTL: CHAR(1) NULL
 HEAD: CHAR(1) NULL
 CNTL_CAT: CHAR(2) NULL
 MISSION_TYPE: CHAR(5) NULL

Country_Codes

FIC: VARCHAR2(3) NOT NULL
 REG: VARCHAR2(2) NULL
 Country: VARCHAR2(30) NULL
 FIPS: VARCHAR2(2) NULL

AREA_CODE_EXCHANGES

Last_Digit: NUMBER(1) NOT NULL
 Area_Code: NUMBER(3) NOT NULL
 Exchange: NUMBER(3) NOT NULL

PH

<p>🔍 ALIAS: VARCHAR2(32) NOT NULL</p> <p>FULLNAME: VARCHAR2(100) NULL EMAILADDRESS: VARCHAR2(200) NULL VERIFYDATE: DATE NULL NICKNAMES: VARCHAR2(200) NULL ADDRESS: VARCHAR2(500) NULL ICD: VARCHAR2(200) NULL PHONE: VARCHAR2(100) NULL FAX: VARCHAR2(100) NULL PAGER: VARCHAR2(100) NULL TITLE: VARCHAR2(200) NULL MISCELLANEOUS: VARCHAR2(500) NULL P_sn: VARCHAR2(30) NULL P_givenName: VARCHAR2(20) NULL P_middleName: VARCHAR2(20) NULL P_generationQualifier: VARCHAR2(4) NULL P_nihNickName: VARCHAR2(200) NULL P_mail: VARCHAR2(200) NULL P_o: VARCHAR2(10) NULL P_buildingName: VARCHAR2(40) NULL P_roomNumber: VARCHAR2(10) NULL P_telephoneNumber: VARCHAR2(25) NULL P_facsimileTelephoneNumber: VARCHAR2(25) NULL P_pagerTelephoneNumber: VARCHAR2(25) NULL P_nihEmailNickname: VARCHAR2(200) NULL UNIQUEIDENTIFIER: CHARACTER(10) NULL BINIT_WT: NUMBER NULL LINKTIMESTAMP: DATE NULL</p>

TELECOM

<p>🔍 LINE_NUM: NUMBER NOT NULL</p> <p>FULLNAME: VARCHAR2(50) NULL SSN: VARCHAR2(20) NULL TITLE: VARCHAR2(10) NULL CAMPUS_PHONE_NUMBER: VARCHAR2(25) NULL ORGANIZATION: VARCHAR2(40) NULL BUILDING: VARCHAR2(20) NULL ROOM: VARCHAR2(10) NULL MAIL_STOP_CODE: VARCHAR2(6) NULL CAMPUS_FAX: VARCHAR2(25) NULL P_GIVENNAME: VARCHAR2(20) NULL P_SN: VARCHAR2(30) NULL P_MIDDLENAME: VARCHAR2(20) NULL P_PERSONALTITLE: VARCHAR2(10) NULL P_TELEPHONENUMBER: VARCHAR2(25) NULL P_FACSIMILETELEPHONENUMBER: VARCHAR2(25) NULL P_NIHSSN: CHAR(9) NULL P_BUILDINGNAME: VARCHAR2(20) NULL P_ROOMNUMBER: VARCHAR2(10) NULL P_O: VARCHAR2(10) NULL P_NIHMMAILSTOP: VARCHAR2(6) NULL UNIQUEIDENTIFIER: CHAR(10) NULL BINIT_WT: NUMBER NULL LINKTIMESTAMP: DATE NULL UPDATETIMESTAMP: DATE NULL P_GENERATIONQUALIFIER: VARCHAR2(4) NULL P_NIHNICKNAME: VARCHAR2(20) NULL</p>

FPS

<p>LASTUPDATE: DATE NULL FNAME: VARCHAR2(13) NULL LNAME: VARCHAR2(19) NULL SSN: CHAR(9) NOT NULL INST: VARCHAR2(12) NULL ADDRESS1: VARCHAR2(30) NULL ADDRESS2: VARCHAR2(30) NULL CITY: VARCHAR2(18) NULL STATE: VARCHAR2(2) NULL PROV: VARCHAR2(15) NULL PO_CODE: VARCHAR2(9) NULL AOUNTRY: VARCHAR2(15) NULL ZIP: CHAR(9) NULL FADDRESS1: VARCHAR2(30) NULL FADDRESS2: VARCHAR2(30) NULL FADDRESS3: VARCHAR2(25) NULL PROVINCE: VARCHAR2(15) NULL POST_CODE: VARCHAR2(9) NULL FOUNTRY: VARCHAR2(15) NULL COUNTRY: VARCHAR2(15) NULL CCODE: VARCHAR2(2) NULL RSTATUS: VARCHAR2(2) NULL EXEMPTCODE: VARCHAR2(1) NULL NAME: VARCHAR2(30) NULL GRANTNO: CHAR(10) NULL EMPNO: VARCHAR2(9) NOT NULL UNIQUEIDENTIFIER: VARCHAR2(10) NULL GRANT_CASESIX_NO: CHAR(6) NULL BINIT_WT: NUMBER NULL LINKTIMESTAMP: DATE NULL PROGRAM_CODE: CHAR(2) NULL CAN: CHAR(7) NULL P_GIVENNAME: VARCHAR2(48) NULL P_IC: VARCHAR2(6) NULL P_SN: VARCHAR2(48) NULL P_MIDDLENAME: VARCHAR2(48) NULL P_GENERATIONQUALIFIER: VARCHAR2(5) NULL P_NIHHOMEHOUSENO: VARCHAR2(6) NULL P_NIHHOMESTREETPREFIX: VARCHAR2(5) NULL P_NIHHOMESTREETNAME: VARCHAR2(20) NULL P_NIHHOME2NDSTREETNAME: VARCHAR2(10) NULL P_NIHHOMESTRUCTDESC: VARCHAR2(4) NULL P_NIHHOMESTRUCTID: VARCHAR2(5) NULL P_NIHHOMERRDESC: VARCHAR2(3) NULL P_NIHHOMERRID: VARCHAR2(5) NULL P_NIHHOMEBOXDESC: VARCHAR2(4) NULL P_NIHHOMEBOXID: VARCHAR2(5) NULL</p>

STATES

<p>STATEIDENTIFIER: NUMBER NULL STATEABBREV: CHAR(2) NULL STATENAME: VARCHAR2(15) NULL</p>
--

CHANGELOG

<p>UNIQUEIDENTIFIER: CHAR(10) NULL Name: VARCHAR2(128) NULL MODIFIERSUID: CHAR(10) NULL MODIFIERSACCESSLEVEL: CHAR(1) NULL AO_UID: CHAR(10) NULL UPDATESTATUS: CHAR(1) NULL SUBMITTIMESTAMP: DATE NULL EXECUTEDTIMESTAMP: DATE NULL SQLSTATEMENT: VARCHAR2(2000) NULL DESCRIPTION: VARCHAR2(1024) NULL EMAILMESSAGE: VARCHAR2(2000) NULL EXECUTEONDATE: DATE NULL LOGENTRYIDENTIFIER: NUMBER(10) NULL</p>

PAID

<p>PAID_UID: CHARACTER(9) NOT NULL</p> <p>LASTNAME: VARCHAR2(30) NULL FIRSTNAME: VARCHAR2(20) NULL MIDDLEINITIAL: VARCHAR2(1) NULL BUILDING: VARCHAR2(40) NULL ROOM: VARCHAR2(10) NULL AREACODE: CHARACTER(3) NULL EXCHANGE: CHARACTER(3) NULL PHONE: CHARACTER(4) NULL HOMEADDR: VARCHAR2(30) NULL HOMECITY: VARCHAR2(20) NULL HOMESTATE: CHAR(2) NULL HOMEZIP: CHARACTER(5) NULL ZIPFOUR: CHARACTER(4) NULL IC: VARCHAR2(10) NULL EMPLOYEESTATUS: VARCHAR2(1) NULL DATEOFENTRY: DATE NULL ACTIVESTATUS: VARCHAR2(1) NULL FARE: VARCHAR2(1) NULL BINIT_WT: NUMBER NULL UNIQUEIDENTIFIER: VARCHAR2(10) NULL LINKTIMESTAMP: DATE NULL P_GENERATIONQUALIFIER: VARCHAR2(6) NULL P_GIVENNAME: VARCHAR2(50) NULL P_MIDDLENAME: VARCHAR2(60) NULL P_SN: VARCHAR2(50) NULL P_NIHNIKNICKNAME: VARCHAR2(20) NULL P_buildingName: VARCHAR2(40) NULL P_roomNumber: VARCHAR2(10) NULL P_telephoneNumber: VARCHAR2(25) NULL P_nihHomeHouseNo: VARCHAR2(6) NULL P_nihHomeStreetPrefix: VARCHAR2(5) NULL P_nihHomeStreetName: VARCHAR2(20) NULL P_nihHome2ndStreetName: VARCHAR2(10) NULL P_nihHomeStructDesc: VARCHAR2(4) NULL P_nihHomeStructId: VARCHAR2(5) NULL P_nihHomeRRDesc: VARCHAR2(3) NULL P_nihHomeRRId: VARCHAR2(5) NULL P_nihHomeBoxDesc: VARCHAR2(4) NULL P_nihHomeBoxId: VARCHAR2(5) NULL P_o: VARCHAR2(10) NULL P_organizationalStatus: VARCHAR2(1) NULL</p>
--

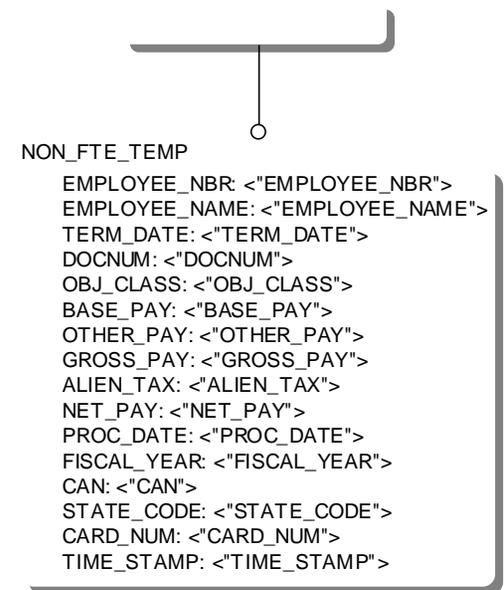
LogLock

<p>LocTime: DATE NULL</p>

ITAS_Info_Master

<p>indv_ssn_id: CHAR(80) NULL eff_date: VARCHAR2(80) NULL indv_last_name: VARCHAR2(80) NULL indv_frst_name: VARCHAR2(80) NULL indv_mid_init: VARCHAR2(80) NULL lan_id: VARCHAR2(80) NULL emai_logn_id: VARCHAR2(80) NULL org_code: VARCHAR2(80) NULL indv_sts_code: VARCHAR2(80) NULL pay_plan_code: VARCHAR2(80) NULL posn_titl_txt: VARCHAR2(80) NULL appt_type_code: VARCHAR2(80) NULL pay_base_rate_symb: VARCHAR2(80) NULL indv_accr_sl_flag: CHAR(1) NULL indv_flex_flag: CHAR(1) NULL indv_pool_flag: CHAR(1) NULL tmkp_flag: CHAR(1) NULL ao_flag: CHAR(1) NULL spvr_flag: CHAR(1) NULL indv_strf_date: VARCHAR2(80) NULL indv_sep_date: VARCHAR2(80) NULL indv_tour_hrs: VARCHAR2(80) NULL last_updt_tmstp: VARCHAR2(80) NULL last_updt_pgm: VARCHAR2(80) NULL last_updt_lan_id: VARCHAR2(80) NULL lv_cat_code: VARCHAR2(80) NULL flsa_flag: CHAR(1) NULL serv_code: VARCHAR2(80) NULL pay_off_flag: VARCHAR2(80) NULL pay_btch_rqst_flag: CHAR(1) NULL hour_rate_dol: VARCHAR2(80) NULL end_date: VARCHAR2(80) NULL al_max_carr_hrs: VARCHAR2(80) NULL schd_type: VARCHAR2(80) NULL dly_type: VARCHAR2(80) NULL shift_code: VARCHAR2(80) NULL mil_sts: VARCHAR2(80) NULL pay_plan_type_code: VARCHAR2(80) NULL spd_rate: VARCHAR2(80) NULL serv_cmp_date: VARCHAR2(80) NULL glbl_post_flag: VARCHAR2(80) NULL mass_aprv_flag: CHAR(1) NULL pay_flag: CHAR(1) NULL fmla_flag: CHAR(1) NULL can_code: VARCHAR2(80) NULL</p>

NON_FTE_TEMP@OGDD



NON_FTE_TEMP

<p>EMPLOYEE_NBR: <"EMPLOYEE_NBR"> EMPLOYEE_NAME: <"EMPLOYEE_NAME"> TERM_DATE: <"TERM_DATE"> DOCNUM: <"DOCNUM"> OBJ_CLASS: <"OBJ_CLASS"> BASE_PAY: <"BASE_PAY"> OTHER_PAY: <"OTHER_PAY"> GROSS_PAY: <"GROSS_PAY"> ALIEN_TAX: <"ALIEN_TAX"> NET_PAY: <"NET_PAY"> PROC_DATE: <"PROC_DATE"> FISCAL_YEAR: <"FISCAL_YEAR"> CAN: <"CAN"> STATE_CODE: <"STATE_CODE"> CARD_NUM: <"CARD_NUM"> TIME_STAMP: <"TIME_STAMP"></p>

ITAS

<p>SSN: CHAR(80) NOT NULL</p> <p>LASTNAME: VARCHAR2(80) NULL FIRSTNAME: VARCHAR2(80) NULL MIDDLEINITIAL: VARCHAR2(80) NULL LANID: VARCHAR2(80) NULL EMAILLOGINID: VARCHAR2(80) NULL ORGCODE: VARCHAR2(80) NULL POSITIONTITLE: VARCHAR2(80) NULL AO: CHAR(1) NULL SUPERVISOR: CHAR(1) NULL PAYOFF: VARCHAR2(80) NULL CAN: VARCHAR2(80) NULL IC: VARCHAR2(20) NULL P_GIVENNAME: VARCHAR2(50) NULL P_MIDDLENAME: VARCHAR2(60) NULL P_SN: VARCHAR2(50) NULL P_GENERATIONQUALIFIER: VARCHAR2(5) NULL IGNORE_ENTRY: CHARACTER(1) NULL UNIQUEIDENTIFIER: VARCHAR2(10) NULL BINIT_WT: NUMBER NULL LINKTIMESTAMP: DATE NULL</p>

NON_FTE_TEMP_M

EMPLOYEE_NBR: CHAR(9) NOT NULL
 EMPLOYEE_NAME: CHAR(22) NOT NULL
 TERM_DATE: DATE NOT NULL
 DOCNUM: CHAR(13) NOT NULL
 OBJ_CLASS: CHAR(4) NOT NULL
 BASE_PAY: NUMBER(7,2) NULL
 OTHER_PAY: NUMBER(6,2) NULL
 GROSS_PAY: NUMBER(7,2) NOT NULL
 ALIEN_TAX: NUMBER(5,2) NULL
 NET_PAY: NUMBER(7,2) NULL
 PROC_DATE: DATE NOT NULL
 FISCAL_YEAR: CHAR(4) NOT NULL
 CAN: CHAR(7) NOT NULL
 STATE_CODE: CHAR(2) NULL
 CARD_NUM: CHAR(1) NULL
 TIME_STAMP: CHAR(26) NULL

FPS_M

LASTUPDATE: DATE NULL
 EMPNO: CHAR(9) NOT NULL
 FNAME: CHAR(13) NULL
 LNAME: CHAR(19) NULL
 SSN: CHAR(9) NOT NULL
 INST: CHAR(12) NULL
 GRANTNO: CHAR(10) NULL
 ADDRESS1: CHAR(30) NULL
 ADDRESS2: CHAR(30) NULL
 CITY: CHAR(18) NULL
 STATE: CHAR(2) NULL
 ZIP: CHAR(9) NULL
 PROV: CHAR(15) NULL
 PO_CODE: CHAR(9) NULL
 AOUNTRY: CHAR(15) NULL
 FADDRESS1: CHAR(30) NULL
 FADDRESS2: CHAR(30) NULL
 FADDRESS3: CHAR(25) NULL
 PROVINCE: CHAR(15) NULL
 POST_CODE: CHAR(9) NULL
 FCOUNTRY: CHAR(15) NULL
 COUNTRY: CHAR(15) NULL
 CCODE: CHAR(2) NULL
 RSTATUS: CHAR(2) NULL
 EXEMPTCODE: CHAR(1) NULL
 NAME: CHAR(30) NULL
 GRANT_CASESIX_NO: CHAR(6) NULL

ROSETTASTONE_LINKS

DIR_UID: CHAR(10) NULL
 TEL_LINE_NUM: NUMBER NULL
 DIR_SSN: CHAR(9) NULL
 TEL_SSN: CHAR(9) NULL
 DIR_SN: VARCHAR2(30) NULL
 TEL_SN: VARCHAR2(30) NULL
 DIR_GIVENNAME: VARCHAR2(20) NULL
 TEL_GIVENNAME: VARCHAR2(20) NULL
 DIR_MIDDLENAME: VARCHAR2(20) NULL
 TEL_MIDDLENAME: VARCHAR2(20) NULL
 DIR_IC: VARCHAR2(10) NULL
 TEL_IC: VARCHAR2(10) NULL
 LINKCODE: CHAR(2) NULL
 BINIT_WT: NUMBER NULL
 LINKTIMESTAMP: DATE NULL

Appendix B “Fast Track” Attribute Descriptions

NOTE: An attribute name has the prefix **nih** when it does not match an X.500, LIPS, or LDAP standard attribute name. This is to avoid conflicts with future standard name usage.

Table B-1 Abbreviations for Data Sources

<i>Symbol</i>	<i>Mnemonic</i>	<i>Description</i>
A	AO	Administrative Officer for owner of entry
B	PAID	Parking/ID Badge/Transhare DB
E	PH	NIH Email Directory and Forwarding Service
F	FPS	Fellowship Payment System
H	HRDB	NIH Human Resources Database
I	Internal	Anyone on NIH internal subnet
J	JEFIC	J. E. Fogarty International Center DB
O	OWN	Owner of entry (individual identified by entry). Not implemented for “Fast Track”.
P	TELCOM	NIH Telecommunications DB (Phone)
S	FTRDB	Fast Track Relational Database System
T	ITAS	Integrated Time and Attendance System
Y	ANY	Anyone

Table B-2 Private Individual Identifying Information Associated with NIH UIDs¹⁰

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access¹¹</i>
nihSSN	permanent or temporary social security number (<i>ddd-dd-dddd</i>)	N	N	AFHP	A	P	A
nihDateOfBirth	Date of birth (<i>yyyy-mm-dd</i>)	N	N	AH	A		A
nihCityOfBirth	City of birth	N	N	AJ	A		A

¹⁰ Attributes not published in directory.

¹¹ Read accesses to these attributes are logged.

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access¹¹</i>
nihStateOfBirth	State or province of birth	N	N	AJ	A		A
nihCountryOfBirth	Country of birth (ISO 3166 code via validation table)	N	N	AJ	A		A
nihAliasGivenName	Other given names associated with uniqueIdentifier	N	Y	ABEFHP	A		A
nihAliasMiddleName	Other middle names (or initials) associated with uniqueIdentifier	N	Y	ABEFHP	A		A
nihAliaseSn	Other surnames associated with uniqueIdentifier	N	Y	ABEFHP	A		A
nihMothersSurname	Mother's maiden surname (currently unused)	N	N	A	A		A
nihGender	M F; M= male; F= female	N	N	AHJ	A		A

Table B-3 Private Home and Personal Locator Attributes¹²

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
homePhone	Home telephone number in full international format	N	N	AJ	A		AO
homeFax	Home fax number in full international format	N	N	A	A		AO
homePostalAddress	Home postal address (<i>street-address, city, state, postal-code</i>) (RFC 2252 LDAPv3 postal address syntax, limited to 6 lines of 30 characters each)	N	N	ABFH	A		AO
nihHomeHouseNo	House number from homePostalAddress	N	N	S			
nihHomeStreetPrefix	Street prefix from homePostalAddress	N	N	S			
nihHomeStreetName	Street name from homePostalAddress	N	N	S			
nihHome2ndStreetName	2 nd street name from homePostalAddress	N	N	S			
nihHomeStreetDirectionSfx	Street directional suffix from homePostalAddress	N	N	S			
nihHomeStructDesc	Structure description from homePostalAddress	N	N	S			
nihHomeStructId	Structure ID from homePostalAddress	N	N	S			
nihHomeRRDesc	Rural route description from homePostalAddress	N	N	S			
nihHomeRRId	Rural route ID from homePostalAddress	N	N	S			
nihHomeBoxDesc	Box description from homePostalAddress	N	N	S			
nihHomeBoxID	Box ID from homePostalAddress	N	N	S			
personalMobile	Personal mobile telephone number in full	N	N	A	A		AO

¹² Protected from public access in directory.

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
	international format						
personalPager	Personal pager number in full international format	N	N	A	A		AO
nihHomeMail	Personal email address	N	N	A	A		AO
nihEmergencyContactCn	Common name of emergency contact	N	N	A	A		AO
nihEmergencyContactPhone	Telephone number of emergency contact in full international format	N	N	A	A		AO

Table B-4 Public Labeling Attributes

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
cn ¹³	(common name) System-generated from givenName , middleName , and sn . Values both with and without middleName generated if middleName attribute exists. Other values may be added by <i>Update From</i> sources.	Y	Y	S	A	B	Y
generationQualifier	e.g. Jr, III from validation table	N	N	ABHJP	A	BP	Y
givenName	First name	Y	N	ABFHJP	A	BP	Y
initials	Initial letters derived from givenName and middleName	N	N	S			Y
personalTitle	e.g. Mr., Dr. from validation table	N	N	ABJP	A	BP	Y
uniqueIdentifier	Assigned by system	Y	N	S			Y
middleName	Middle name or initial	N	N	ABHJP	A	BP	Y
sn	(surname) Last name	Y	N	ABFHJP	A	BP	Y
nihSuffixQualifier	e.g. MD, PhD from validation table	N	N	AJP	A	BP	Y
description	Free-form, multi-line	N	N	A	AO		Y
nihEmailNickname	Nicknames from NIH Email Directory and Forwarding Service	N	Y	E	AO		Y
nihNickname	Nicknames (givenNames only)	N	Y	ABFHJP	AO	BP	Y
businessCategory	Terms that identify a person's business, technical, special interest, or functions, e.g. "scientist", "molecular biology" (free-form)	N	Y	A	A		Y
jpegPhoto	Full size ID photo in jpeg binary format	N	N	-	B		I(??)

¹³ nihCn* fields in FTRDB.

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
thumbnailPhoto	Thumbnail ID photo in jpeg binary format	N	N	-	B		I(??)

Table B-5 Public Organizational Attributes

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
title	Organizational title, e.g., Chief, CFB	N	Y	AH	A ¹⁴		Y
nihSeriesTitle	Occupational code corresponding to the GS Series	N	N	H	H		Y
secretary	timekeeper UID (currently unused)	N	N	AT	A		I
manager	supervisor UID (currently unused)	N	N	ATJ	A		Y
o	Institute or Center (IC) abbreviation; "NIH" if no IC	Y	N	ABFHJP		BP	Y
organizationalStatus	C F G N V ; C =contract; F =fellow; G =guest; N =NIH FTE; V =volunteer; T =tenant	Y ¹⁵	N	ABFHJ	A		Y
ou	Name and abbreviation of organization unit, generated by system from nihSAC	N	Y	S			Y
nihOrgPath	Organizational path name (e.g. "/NIH/CIT/OCRS/CFB/DSS/"). Determines distinguished name of entry in organizational view.	N	N	S			Y
nihSAC	NIH administrative code of person's ou	N	N	AH	A		Y
nihTelecomOu	Organizational abbreviation used by NIH Telecommunications DB without IC component	N	N	AP	A	P	Y
nihCompanyName	Person's primary employment affiliation if not NIH	N	N	A	A	B	I(??)
nihCompanyPhone	Company telephone number in international format	N	N	A	A	B	I(??)

Table B-6 Public Locator Attributes

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
labeledURI	URL of NIH related WEB site	N	Y	-			Y
mail	Preferred email address	N	N	AE ¹⁶	A		Y

¹⁴ Can be updated when an organizationalRole directory entry is created.

¹⁵ Required for Nih business data; not for directory objectclass definition.

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
nihUniqueMail	Assigned email address @nih.gov	N	N	ET	AET	T(??)	Y
facsimileTelephoneNumber	Office fax number (international format)	N	N	AEP	A	P	I
mobileTelephoneNumber	Office mobile number (international format)	N	N	A	A		I
pagerTelephoneNumber	Office pager number (international format)	N	N	AE			I
telephoneNumber	Office telephone number (international format)	N	N	ABEPJ		PB	Y
buildingName	Office building designator	N	N	ABEPJ	A	PB	Y
houseIdentifier	Same as buildingName . This attribute is not stored in the FTRDB or displayed on any forms—it exists only as a read-only, system generated attribute in the directory.	N	N	S			Y
roomNumber	Room designator for office	N	N	ABEPJ	A	PB	Y
st	(state) State name for office. Generated by system from buildingName .	N	N	S		PB	Y
c	(country) Always “US”	Y	N	S			Y
l	(locality) City name, or other local designator, for office. Generated by system from buildingName .	N	N	S		PB	Y
nihPhysicalAddress	Physical location of office (RFC 2252 LDAPv3 postal address syntax, limited to 6 lines of 30 characters each)	N	N	S			Y
street	Street address and name for office	N	N	S			Y
nihMailstop	NIH mail stop code	N	N	AP	A	P	Y
nihDeliveryAddress	Delivery address for private carriers (e.g., FedEx, UPS);	N	N	A	A		Y
postalAddress	Full USPS address, including street address, city, state, postal code, etc., to which mail can be sent. (RFC 2252 LDAPv3 postal address syntax, limited to 6 lines of 30 characters each)	N	N	A	A		Y
postalCode	USPS ZIP	N	N	A	A		Y

¹⁶ System generated, same as “nihUniqueMail” during Fast Track.

Table B-7 Security Attributes

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
userCertificate	Public Key Certificate	N	Y	S			Y
userPassword	Password to directory	N	N	S	O		

Table B-1 Ancillary Attributes

<i>Attribute</i>	<i>Description</i>	<i>Req</i>	<i>Multi Valued</i>	<i>Initial Source</i>	<i>Update From</i>	<i>Update To</i>	<i>Read Access</i>
creatorsName	UID of administrator creating entry	Y	N	S			
createTimestamp	Time stamp of nihInetOrgPerson creation event	Y	N	S			
modifiersName	UID of last administrator modifying entry	Y	N	S			
modifyTimestamp	Time stamp of last nihInetOrgPerson modify event	Y	N	S			
nihPersonStatus	A I T S ; A=active; I=inactive; T=transferring; S=suspended	N	N	ABFHJ	A	B	A
nihUidQuality	0=not validated; 1=3 rd party; 2=personal contact	Y	N	A			I
nihUidValidator	UID of administrator validating UID	N	N	S			I
nihUidValidationTimestamp	Time stamp of UID validation event	N	N	S			I
nihDirEntryUnlisted	Y N ; Y=unlisted directory entry. For "Fast Track", causes entry not to be present on LDAP server. NOTE: since public directory information for Federal employees is subject to the Freedom of Information Act, an entry cannot be unlisted without justification.	N	N	A	A		AO
nihDirEntryEffectiveDate	Date directory entry becomes active	N	N	A(FHJ??)			I
nihDirEntryExpirationDate	Date directory entry expires	N	N	A(BFHJ??)			I

Appendix C “Fast Track” Database Creation

C.1 Record Linking

1. Create UIDs for HRDB, FPS, and JEFIC after linking to identify overlaps.
 - Nearly all HRDB and FPS records should have SSNs, and should not overlap.
 - JEFIC Visiting Fellows can be linked to FPS on the JEFIC case number. Report Visiting Fellows that cannot be linked for review.
 - JEFIC Visiting Scientists and Associates can be linked to HRDB on name, DoB, and home address. Report Visiting Scientists and Associates that cannot be linked for review.
 - JEFIC “Others” should not overlap HRDB or FPS.

Associate the individual identifying information from these with UIDs, and add to the FTRDB. Report UIDs with duplicate SSNs for review. Select and load the name attributes (`personalTitle`, `givenName`, `middleName`, `sn`, `generationQualifier`, and `nihSuffixQualifier`), `homePhone`, `homePostalAddress`, `organizationalStatus`, `title` (from job series), `o` (authoritative), `ou`, `nihSAC`, `nihDirEntryEffectiveDate`, and `nihDirEntryExpirationDate` attributes with the values found in the most recently modified record. Load work address, work telephone, and manager¹⁷ from JEFIC.

 - Save any other conflicting names as alias attributes (`nihAliasGivenName`, `nihAliasMiddleName`, and `nihAliasSn`) and as `cn` attributes.
2. Link records from ITAS to the FTRDB on SSN and select and load the `nihUniqueMail` attribute.
3. Link records from the Telecommunications database to the FTRDB on SSN.
 - Select and load the name attributes, `telephoneNumber`, `facsimileTelephoneNumber`, `buildingName`, `roomNumber`, `nihTelecomOu`, and `nihMailstop` attributes.
 - Save conflicting names as alias and `cn` attributes.
 - Save conflicting given names as `nihNicknames`.
4. Link records from the PAID database to the FTRDB on name, `organizationalStatus`, home address, organization, work address(?), and work phone(?).
 - Report conflicts in `homePostalAddress` and `organizationalStatus` for review(?).
 - Save conflicting names as alias and `cn` attributes(?).

Select and load the name attributes, `homePostalAddress`, `buildingName`, `roomNumber`, and `telephoneNumber` attributes.

At this point, the FTRDB should contain entries for about 21,000 individuals, with UIDs associated with individual identifying information of good quality. The following steps add entries for individuals with UIDs which will have little or no associated individual identifying information.

5. Add new UIDs for active¹⁸ records in PAID that have not yet been associated with a UID. These should be individuals who are contractors, visitors, and guests, and who are not JEFIC “Others”. Load the name attributes, `homePostalAddress`¹⁹, `organizationalStatus`, `o`, `buildingName`, `roomNumber`, and `telephoneNumber` attributes.
6. Link records from the Telecommunications database to the FTRDB on name, `o`, `nihTelecomOu`, `telephoneNumber`, `buildingName`, `roomNumber`, attributes. Report links with conflicting

¹⁷ Converting JEFIC “sponsor” to “manager” unique identifier still to be tested.

¹⁸ “Active” means individuals with non-expired ID badges. Records are retained in PAID for individuals with expired ID badges, but with non-expired parking permits or TRANSHARE passes. These individuals will be entered in the directory when they renew their expired ID badges.

¹⁹ Home address, or vehicle registration information might serve as weak individual ID information for holders of NIH parking permits.

SSNs for clerical review. Select and load: name, nihTelecomOu, telephoneNumber, buildingName, roomNumber,

7. Link records from the NIH Email Directory and Forwarding Service to the FTRDB on name, o, nihUniqueMail, telephoneNumber, facsimileTelephoneNumber (?), and physical office location(?).

- Save conflicting names as alias and cn attributes.
- Save conflicting given names as nihNicknames.
- Report conflicts in nihUniqueMail for review.

Load nihEmailNickName attribute. Select and load nihUniqueMail, title, facsimileTelephoneNumber, pagerTelephoneNumber, telephoneNumber, buildingName, and roomNumber attributes.

- Do **not** copy the mail attribute from the NIH Email Directory until NIH directory functions as or connects to @nih.gov forwarding service. The mail attribute will be generated automatically by the system from nihUniqueMail.
- Records in the email and telecom directories that cannot be matched will **not** be included in the FTRDB²⁰.
- The record linking operations in the above steps may result in multiple links to a single UID in the FTRDB, indicating possible multiple records for the same individual in the file being linked to the FTRDB, or links from a single record in the file being linked to multiple UIDs in the FTRDB, indicating the possible assignment of multiple UIDs to the same individual. Such multiple links must be resolved manually.
- The record linking operations in the above steps may also result in *borderline links*—i.e. links with a binit weight near zero, indicating that the pair of records cannot be classified either as linked or unlinked. Such cases should be manually reviewed, looking for possible tests or weight changes to incorporate in the linking engine to improve its discriminating power.

C.2 Attribute Selection

When multiple records from different databases are joined or linked to the same UID, conflicting values for the same attribute may result. These conflicts will be resolved and attributes selected for loading as follows:

- Name attributes
 1. HRDB, ITAS, FPS, JEFIC, and PAID are the authoritative sources, in that order of priority, for the following attributes: generationQualifier, givenName, personalTitle, middleName, sn, nihSuffixQualifier.
 2. Different values for givenName, middleName, and sn from other sources are not treated as conflicts, but are saved as nihAliasGivenName, nihAliasMiddleName, nihAliasSn cn, and nihNickName attributes²¹.
- Private Home and Personal Locator Attributes
 - HRDB, FPS, JEFIC, and PAID are the authoritative sources, in that order of priority, for the following attributes: homePhone, homePostalAddress.
- Public Organizational and Locator Attributes

²⁰ We believe that excluding email and telecom directory records that cannot be matched to any record in at least one of the other databases is advantageous because (1) we expect that many or most of these will belong to individuals who are no longer affiliated with NIH, and (2) these individuals, if they are affiliated with NIH, are not subject to registration/update/deregistration by an AO.

²¹ Consider only saving different names if they are significantly different (e.g. have a different SOUNDEX code) to screen out typographical errors.

1. HRDB, FPS, JEFIC, and PAID are the authoritative sources, in that order of priority, for the following attributes: `organizationalStatus`, `o`, and `ou`.
2. Attributes will only be loaded from records with a matching `o` (IC) attribute.
3. Attributes from Telecommunications will be loaded in preference to those from PAID or the NIH Email Directory.
4. Conflicting attributes will be selected from PAID or the NIH Email Directory based on which was more recently modified.

Appendix D Probabilistic Record Linkage

Linkage is the bringing together of information from two database records that relate to the same individual. Calculating the likelihood that the linkage is correct makes the linkage process *probabilistic*.

The degree of certainty that a linkage is correct depends upon the comparisons of available attributes (or fields) of the records, and the outcomes of these comparisons. Generally, *agreement* between the values of an attribute in a pair of records argues in favor of accepting them as a *linked pair*, while *disagreement* of attribute values is characteristic of an *unlinked pair*.

However, agreement of various attributes and values have varying significance. For example:

- It is more likely that two records that agree only on **surname** are linked than two records that agree only on **first name**.
- It is more likely that two records that agree on **surname** = “GORLEN” are linked than two records that agree on **surname** = “SMITH”.
- It is more likely that two records with **surnames** = “SMITH” and “SMITHE” are linked than two records with **surnames** = “SMITH” and “BROWN”.

The odds that a linkage is correct can be calculated by measuring the frequency of the outcomes of a comparison applied to a representative set of linked pairs, and dividing that by the frequency of the outcomes of the same comparison applied to a representative set of unlinked pairs:

$$\text{odds} = \frac{\text{frequency of outcome}(x, y) \text{ among linked pairs}}{\text{frequency of outcome}(x, y) \text{ among unlinked pairs}}$$

where:

- x indicates the attribute and its value on the record from database A
- y indicates the attribute and its value on the record from database B

When multiple comparisons involving various attributes and values are performed on a pair of records, the overall odds of correct linkage are calculated by simply multiplying together the odds of the individual comparisons. However, it is customary to express the odds as a *binit weight*:

$$\text{binit weight} = \log_2 \text{odds}$$

and to then calculate the total binit weight by summing the binit weights of the individual comparisons.

Note that the representative set of linked pairs need not be large (a few hundred is sufficient to start with), and it need not be perfect. Applying the linkage process generates more linked pairs, which can be added to the set, and the process iterated.

A representative set of unlinked pairs is not required if simple comparisons are used, because the outcome frequencies can be calculated. Care must be taken when performing complicated (and more powerful) comparisons, which involve:

- specific attribute values, e.g. “GORLEN” and “SMITH”
- partial agreement of values, e.g. “SMITH” and “SMITHE”
- comparisons of logically related identifiers, e.g. first names agree **given that** surnames agree

Alternatively, one can estimate the outcome frequencies using any of several procedures, such as the Expectation-Maximization (EM) algorithm.

Finally, there are a multitude of tips and tricks for handling missing values and comparing surnames, initials, given names, dates and places of birth, and geographic attributes.

D.1 Record Linking Example

As a simple example of record linking, suppose that four attributes are compared on representative sample sets of linked and unlinked pairs of records from two files, and the outcome frequencies (expressed as percentages) are obtained as shown in Table D-1.

Table D-1 Example Attribute Comparison Outcome Frequencies

Attribute	Outcome	Linked Pairs	Unlinked Pairs	Ratio	Binit Weight
Surname	Agree	96.5%	0.1%	965.0000	9.9
	Disagree	3.5%	99.9%	0.0350	-4.8
Given Name	Agree	79.0%	0.9%	87.7778	6.5
	Disagree	21.0%	99.1%	0.2119	-2.2
Date of Birth	Agree	93.3%	8.3%	11.2410	3.5
	Disagree	6.7%	91.7%	0.0731	-3.8
Place of Birth	Agree	98.1%	11.7%	8.3846	3.1
	Disagree	1.9%	88.3%	0.0215	-5.5

Now suppose that a pair of records from the two files are compared, and they *agree* on **surname**, **date of birth**, and **place of birth**, but *disagree* on **given name**. Then the total binit weight of the link for the pair is calculated by adding the **bold face** binit weights for the individual outcomes from Table D-1:

$$9.9 + (-2.2) + 3.5 + 3.1 = 14.3$$

This binit weight indicates odds of about $2^{14.3} \cong 10,000$ to 1 that the pair should be linked

D.2 Record Linking References

1. Gill, L. E., and Baldwin, J. A. (1987), "Methods and technology of record linkage: some practical considerations" in J. Baldwin, E. D. Acheson, and W. Graham (ed.) Textbook of Medical Record Linkage, Oxford: Oxford University Press, 39-54.
2. Newcombe, H. B. (1988), Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business, Oxford: Oxford University Press. Classic book reference. Covers some of the theory and much of the heuristics needed for good record linkage practice. Now out of print.
3. Newcombe, H. B. (1987), "Record linking: the design of efficient systems for linking records into individual and family histories" in J. Baldwin, E. D. Acheson, and W. Graham (ed.) Textbook of Medical Record Linkage, Oxford: Oxford University Press, 15-38.
4. Winkler, W. E. (1994), "Advanced Methods of Record Linkage," American Statistical Association, Proceedings of the Section of Survey Research Methods, 467-472. Describes new theory and algorithms in computer science, operations research, and statistics that were developed at the Census Bureau and used in current Census system. Extends original Jaro string comparator and gives likelihood-based methods for connecting the comparators to the main decision rule of Fellegi and Sunter. Introduces a new assignment algorithm for forcing 1-1 matching that is as fast the benchmark Burchard-Derigs algorithm and uses 1/500 as much storage; is also much faster and uses less storage than the MCF algorithm of Klingman. Gives general theory extending EM ideas of Meng and Rubin (Biometrika 1994) and shows how it is applied in estimating record linkage parameters. Gives method for estimating record linkage error rates that holds in more situations than the Belin-Rubin method, that

does not require a training set as does the Belin-Rubin method, and requires an ad hoc intervention that tends to limit its application to record linkage experts.

5. Winkler, W. E. (1995), "Matching and Record Linkage," in B. G. Cox et al. (ed.) *Business Survey Methods*, New York: J. Wiley, 355-384. Survey article that gives much background about record linkage. Describes available software, list acquisition and preparation, and a large number methods for evaluating the quality of lists and the quality of matching results.

Appendix E “Fast Track” Update Service User Interface

Appendix F Outstanding Questions and Issues

1. Propagate updates from directory update service to ph/CSO? A: No.
2. Changes to attributes in the NIH email directory made directly by users will be overwritten by changes made by AOs. A: Not a problem--see #1.
3. Provide for "unlisted" directory entries? A: Yes. For the "fast track" directory an "unlisted" entry will not be transferred to the LDAP directory server.
4. Can "weak" UIDs be eliminated or upgraded to strong UIDs?
5. What individual identifying information can AOs acquire and enter from NIH employees? From NIH contractors?
6. Can/should we switch NIH UIDs to the ISO/IEC 7812-2 identification card standard?
7. Is it feasible to use S/MIME to secure email notifications from the FTUS to other organizations and systems?
8. Would it be feasible to add SOUNDEX codes for surnames contained in common name attributes added by the entry owner?
9. Add attributes for hair color, eye color, and height?
10. Add digitized written signature image attribute?
11. Allow listing of other than sn, giveName, etc. in phone book?
12. Check with Tom Boyce re: status of tracking contractors in ITAS.
13. Privacy and practicality issues re: collecting mother's maiden name.

Appendix G Wish List

- Handle delegation and work flow for forms 1308-4/5 and 443
- Provide way to edit and display alternate locator information
- Provide locator information to NIH mail room
- Hook up to CIT *Remedy* trouble ticket system
- Provide Automated Clearinghouse (ACH) banking information to ADB
- Interface to Employee Express
- Exchange data with Visual Employee Data System (VEDS) (contact: Netcom)
- Exchange data with future WATSON employee system, which will track personnel actions outside of official systems (contacts: Steve Benowitz, Kevin Murphy, and Mike Pollard (SMC))
- Obtain full organization name for SACs from Scientific Directory when available
- Hook up to Radiation Safety database
- Provide forwarding address and phone for ex-employees